**Instructor:** Cameron Thrash

[thrash@usc.edu](mailto:thrash@usc.edu)

**Office hours:** by request

**Meeting times:** Friday 13:00-15:00, location TBD

**Textbooks:**

*Practical Computing for Biologists*, Haddock & Dunn

*Phylogenomics*, DeSalle & Rosenfeld

**Websites:**

[https://blackboard.usc.ed](https://blackboard.usc.edu)u

<https://hpcc.usc.edu/>

**Course Description**

Contemporary research in the biological sciences has capitalized on recent advances in computational methodologies to address novel questions relating to the ecology and evolution of life on this planet. To effectively contribute to this growing field, the next generation of biologists must have a solid foundation in both classical biology and bioinformatics. This hands-on course will introduce biologists with minimal to no prior command line experience to basic bioinformatics skills, software, and analysis pipelines.

**Prerequisites:** Admission to the MBBO or Ocean Sciences graduate programs or permission from the instructors.

**Specific learning objectives. By the end of this course,** students will be able to:

1. Submit batch jobs to an HPC system.
2. Execute a wide variety of command line operations.
3. Complete data transfer between multiple types of computing resources.
4. Create basic bash scripts.
5. Use several common bioinformatics programs.
6. Troubleshoot utilization of a novel software/analysis pipeline.

**Tentative schedule of lecture topics**. *Note that the schedule is subject to change*. Any schedule changes will be discussed in class and posted on Blackboard.

|  |  |  |  |
| --- | --- | --- | --- |
| **Day** | **Topic** | **Readings, etc.** | **Assignment** |
| 08/21 | Intro to HPC/command line | Software carpentry tutorials  Getting started with USC HPC | 1. Logging + Path Files |
| 08/28 | Editing text files, customizing your work environment, file transfers | Haddock and Dunn Chps 4,5,20 | 2. Linux commands Pair and Share |
| 09/04 | Linux one-liners | Software carpentry Unix + Optional: Regex tutorials  Optional: Haddock & Dunn Ch 2,3 | 3. Piped Commands |
| 09/11 | Shell scripting basics + loops + submitting jobs | Haddock and Dunn Chp 6  USC HPC Running a Job | 4. Bash Scripts |
| 09/18 | Perl/Python libraries; Downloading, viewing, and quality filtering FASTQ files Part I | TBD | 5. Independent QF |
| 09/25 | Quality filtering FASTQ files Part II | Goodwin et al (2016). *Nature Reviews Genetics,* 17: 333-351.  Evolution of DNA Sequencing Methods, Jonathan Eisen: <https://www.youtube.com/watch?v=s9UbA7VyISQ> | 6. Sequencing technologies |
| 10/02 | Genome assembly I: Final project development; conceptual assembly (if time) | Installations | Genome assembly and final stats |
| 10/09 | Genome assembly II: read mapping, QC  Conceptual assembly | Installations | 7+8. Two-week report, work on semester projects |
| 10/16 | Annotation/BLAST/HMMs | DeSalle & Rosenfeld, Chp 5, 7  (Optional) Eddy, S. R. (2004). What is dynamic programming? *Nature Biotechnology*, *22*(7), 909–910. | Work on semester projects |
| 10/23 | Phylogenetics | De Salle & Rosenfeld Chps 6,8 | Work on semester projects |
| 10/30 | Open help time |  | Work on semester projects |
| 11/06 | Open help time |  | Work on semester projects |
| 11/13 | Final Project Presentations |  |  |

**Dealing with challenges**

Making mistakes and running into roadblocks is inherent to the process of learning. By design, this course will challenge you to figure out solutions, not simply give you the path from point A to point Z. The reason for this is that the struggle to overcome whatever challenge you face is where the true learning happens. Therefore, while your first impulse when you encounter a problem will probably be to email me, this will not be met with the type of response you are looking for UNLESS you have done all of the following problem-solving efforts first, in this order:

1. **Think.** Review your commands, inputs, and outputs. See if you can figure out what went wrong. Often it’s simply that a space is missing somewhere, a comma is misplaced, or you’re using a ‘ instead of a `. Try to find this yourself before bothering anyone else.
2. **Consult the internet.** The people who have developed and use the programs you are learning have created a massive number of online resources. Often googling your error message will allow you to find the problem. Creative google searching can usually do the rest.
3. **Consult your peers.** Is everyone in the class having the same problem, or has someone else discovered the solution? While this may seem like the same thing as asking the professor, asking your classmates fosters peer-to-peer instruction, which reinforces concepts for those who get the opportunity to teach their solution and gives a different perspective to those who are seeking answers.
4. **Consult Dr. Thrash.** If you’re still having problems, by all means contact the instructor or set up an appointment for office hours. Solutions are sometimes very simple but obscure.

**Staying Organized**

Part of any good computational biology workflow is keeping your inputs and outputs organized, and all your processes and the contents of each file and directory documented. This not only allows someone else to understand and reproduce your work, but prevents you from forgetting the valuable steps you took to produce your work as well. It’s a horrible feeling to enter a directory a year after working on a project and not remember the contents of the files or how they were created. Throughout the semester, we will utilize a common core set of organizational procedures to facilitate keeping organized. Each week will have:

* A separate directory in your home directory where you will store inputs, outputs, and possibly include subdirectories (1 pt)
* README file that documents the contents of the directories/subdirectories (4 pts)
* Report, described below (10 pts)

All three of these elements will be instrumental in your Weekly Assignment grade, based on the point totals indicated.

(You may find that when you branch out on your own, a different system may suit you. Regardless, it is important to leave a transparent trail of all your work so that it can be recreated at any point in the future. The point here is simply to enforce good practices in computational biology, and we have to pick one system in advance.)

**Reports**

Each week you will be completing a series of tasks using a tool or set of tools. As part of your assignments you need to include a short written report with the elements below. The point of the report is to document your computational workflow so that it is REPRODUCIBLE from scratch. This means they can either include the commands themselves, the scripts they used, or both, but one must be able to recreate the results based on the information provided. If the report includes only text, create it with a text editor (e.g. nano) and save as ~/<working directory>/report.txt. If it includes graphics, save as a .docx or .pdf file, and upload to ~/<working directory>/report.docx (or .pdf).

1. Name(s), date
2. General (1-line) summary of objective(s) and purpose(s)
3. Working directory
4. Programs used, including basic scripts, and relevant reference(s)
5. Commands, inputs, outputs and results/evaluation of output for NEW operations.\*

* Organize in sections consistent with the assignment workflow
* Include specific file names
* For batch jobs, you can point to a script if it contains the commands needed, or you can include the command in the report.
* For repeating tasks, only detail the first instance, then indicate that this was repeated and note variation in input/output file names. Similarly, output only has to be shown for a first instance.
* \*For operations you are repeating from previous assignments (blastp, muscle, etc.), you may simply reference a previous report, but you need to be specific enough that one could find the correct command and repeat it.

1. Personal reflection. What did you learn in addition to using the assigned tool(s)? What would you do differently? What are you still confused about?

**HPC Resources & Software**

Students must have a personal computer and an account on the USC HPC system to participate in this class. You have been assigned an account within our course allocation which can be accessed at /home/rcf-proj/jct1/<your USC username>/ and we will review how to login during the first class. Two-factor authentication is needed for login, so please bring DUO-enabled mobile device to class.

Students working on a PC will also need to install an SSH client – we recommend PuTTY: <https://www.putty.org/>

Please also download an FTP client - we recommend WinSCP <https://winscp.net/eng/index.php> for PC’s or FileZilla for Mac/PC: <https://filezilla-project.org/>

All students should also have a text editor installed. Those working on a PC can use Notepad. TextEdit is recommended for Mac users: <https://www.barebones.com/products/textwrangler/>

We also recommend applying for a personal HPC account in parallel to facilitate application of course content to your specific dissertation research. Accounts are free and your supervisor may already have an allocation – if so, request that they add you as a user. If you do not already have access to an allocation, please apply for one here:

<https://hpcc.usc.edu/support/accounts/applying-for-a-hpcc-account/>

**Additional Resources**

Learning regex:<http://www.regexr.com>

Linux cheat sheet:<http://peoplesofttutorial.com/learn-basic-linux-commands-using-linux-cheat-sheet/>

Rosalind programming training:<http://rosalind.info/problems/locations/>

Software Carpentry training:<http://software-carpentry.org/index.html>

Elements of Bioinformatics:<http://elements.eaglegenomics.com>

**Course Policies**

COMMUNICATION: Email is the preferred form of communication with the instructor outside of class. Responses can be expected within 24-48 hours, though there may be an additional delay over weekends/holidays. If an in-person meeting is desired, please email a request to set up a meeting time.

GRADES: Any document associated with grading may be photocopied by the instructional staff.

The final letter grade will be determined by the total number of points as follows.

|  |  |  |
| --- | --- | --- |
|  | Points possible Percent | |
| Weekly Assignments (15 points ea.) | 120 | 40% |
| In-class Contributions (5 points ea.) | 50 | 17% |
| Final Project Report + Presentation | 130 | 43% |
| **TOTAL** | 300 | 100% |

WEEKLY ASSIGNMENTS: Exercises to help you develop your skills will be conducted during each class period. Some of these may involve additional work outside of class.

IN-CLASS CONTRIBUTIONS: Good attendance is necessary but not sufficientto earn full marks for this component of the course grade. Students who earn top scores will participate in class exercises and discussions actively, thus helping lead their peers to think critically and analytically. This requires students to chime in, ask questions, and makr relevant comments in response to the questions of others. It also necessitates full participation and completion of in-class exercises and problem sets.

FINAL PROJECT: You will be responsible for utilizing some combination of the tools and skills learned in the course to complete real-world analyses. This can include analyses of data connected to your dissertation research, or if no such data is available, that from a publically available repository.

PLANNED ABSENCES: Requests for absences should be made by email to the Instructors at least 2 weeks in advance and preferably as early in the semester as possible. If the absence is approved a reasonable accommodation will be provided, which may include coordinating remote participation or a make-up assignment.

### **Statement on Academic Conduct and Support Systems**

**Academic Conduct:**

Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, “Behavior Violating University Standards” [policy.usc.edu/scampus-part-b](https://policy.usc.edu/scampus-part-b/). Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, [policy.usc.edu/scientific-misconduct](http://policy.usc.edu/scientific-misconduct).

**Support Systems:**

*Student Health Counseling Services - (213) 740-7711 – 24/7 on call*

[engemannshc.usc.edu/counseling](https://engemannshc.usc.edu/counseling/)

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

*National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call*

[suicidepreventionlifeline.org](http://www.suicidepreventionlifeline.org/)

Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

*Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-4900 – 24/7 on call*

[engemannshc.usc.edu/rsvp](https://engemannshc.usc.edu/rsvp/)

Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

*Office of Equity and Diversity (OED) | Title IX - (213) 740-5086*

[equity.usc.edu](https://equity.usc.edu/), [titleix.usc.edu](http://titleix.usc.edu)

Information about how to get help or help a survivor of harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants. The university prohibits discrimination or harassment based on the following protected characteristics: race, color, national origin, ancestry, religion, sex, gender, gender identity, gender expression, sexual orientation, age, physical disability, medical condition, mental disability, marital status, pregnancy, veteran status, genetic information, and any other characteristic which may be specified in applicable laws and governmental regulations.

*Bias Assessment Response and Support - (213) 740-2421*

[studentaffairs.usc.edu/bias-assessment-response-support](https://studentaffairs.usc.edu/bias-assessment-response-support/)

Avenue to report incidents of bias, hate crimes, and microaggressions for appropriate investigation and response.

*The Office of Disability Services and Programs - (213) 740-0776*

[dsp.usc.edu](http://dsp.usc.edu/)

Support and accommodations for students with disabilities. Services include assistance in providing readers/notetakers/interpreters, special accommodations for test taking needs, assistance with architectural barriers, assistive technology, and support for individual needs.

*USC Support and Advocacy - (213) 821-4710*

[studentaffairs.usc.edu/ssa](https://studentaffairs.usc.edu/ssa/)

Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

*Diversity at USC - (213) 740-2101*

[diversity.usc.edu](https://diversity.usc.edu/)

Information on events, programs and training, the Provost’s Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

*USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call*

[dps.usc.edu](http://dps.usc.edu/), [emergency.usc.edu](http://emergency.usc.edu/)

Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

*USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-120 – 24/7 on call*

[dps.usc.edu](http://dps.usc.edu/)

Non-emergency assistance or information.